Proposal for a Unified "Flux" N-tuple Format.

Robert Hatcher
FNAL/CD

April 3, 2012

1 Statement of Purpose

The FNAL neutrino experiments (MINOS, MINERνA, NOνA, ArgoNeut, MicroBooNe, LBNE) all have similar needs for simulations of the beamlines. Each of the NuMI, Booster and LBNE beamlines send protons into their respective targets, producing secondaries that decay to neutrinos; by keeping sufficient information those decays can be re-evaluated for different detector locations by event generators such as GENIE.

Various groups have used different tools to model the physics and geometry of the beamlines. These include combinations of GEANT3, GEANT4 and FLUKA. Unfortunately, over time, these simulations have comt to have incompatible variants in the structure of their outputs. Some of these differences include a change of basic types, capitalization of the leaf element names, changes in array sizes, and additions of variables. This makes it more difficult for the different groups to make comparisons and to use common tools. GENIE's flux interface `GNuMIFlux` must support all the variants. This gets more difficult as individual, incompatible twists are introduced.

I am proposing that a single new format be defined and that all beamline simulations be modified to fill that format. The new structure should be an intelligent union of all the core parts and individual extensions. If a particular simulation doesn't generate or wish to store a non-essential element then they would flag it as unfilled. Additionally provisions would be made to use C++ STL `vector`s rather than fixed array sizes to allow for more flexibility and less waste. A scheme for proprietary (temporary) extensions should also be designed in to allow open-ended studies without the need for significant code changes. Below, I attempt to identify existing `Branch`es in the various `TTree`s and show their existing status and the new proposal.

It would also be useful to introduce a mechanism to record in the file some metadata that applies to the file as a whole. This includes total protons-on-target (rather than trying to infer it from the range of `evtno`); the actual detector locations used for "near" and "far"; and statements about the tools used to generate the file (e.g. flugg, geant4, etc. and build version).

This might also be a good time to rename the GENIE `GNuMIFlux` class to avoid prejudice against Booster and LBNE beam simulations; a `typedef` could be used to retain backward compatibility. The `GNuMIFluxPassThroughInfo` class would migrate to be identical in form to this new layout and undergo a renaming.

Thanks to Alex Himmel for producing MINOS-DocDB-6316 from whence I stole a lot of tables to serve as a starting point for this document.

2 Primary Ntuple

2.1 general characterisics

The primary ntuple holds entries representing decays that produced neutrinos with one entry for every neutrino recorded (generally with some importance weight). It is possible for the same initial proton to produce more than one entry (i.e. the same `evtno` might appear more than once).

The MINERνA variant of the `g4numi` layout appears to only add new branch elements which are discussed in Table 7.

simulation	base program(s)	tree name	capitalization	char limit
gnumi	geant3	h10	first char, sometimes	8 char
flugg	fluka + geant4	h10	follows gnumi	8 char
g4numi	geant4	nudata	studly, e.g. `NdxdzNear`	none
lbne	geant4	nudata	follows g4numi	none
—	— all —	dk2nu	all lower case	none

Table 1: General properties of the ntuples.

At this time the format of any given ntuple file must be guessed from a combination of the file and tree names. By choosing a new unique tree name (e.g. `dk2nu`) for the new `TTree` format it can be easily identified; alternative suggestions for this name are welcome. I propose that branch element names for the new format are entirely lower case for ease of rememberence and typing. Also no artificial name cutoffs should be imposed (i.e. `ndxdznear` rather than `NdxdzNea`).

Each sub-section below tabulates a number of branch elements, gives their type for each `TTree` variant and a general description. These are grouped only for convenience and it is the aggregate that makes up the `TTree` structure.

Notes:

1. $\hat{z}$ is beam direction, centerline axis
2. energy & momentum are in GeV [allow to flag for MeV with `flagbits`? ‡]
3. distances in cm [allow to flag flag for m or mm with `flagbits`? ‡]
4. particle codes Geant3 [change default to PDG, flag old with `flagbits`? ‡]
5. branch types: I=integer; F=float; D=double; TS=`TString`; s=STL `string`
6. $[n]$ = fixed size array; <>= STL vector
7. if type is ? then either type conflict or unknown whether final ntuple needs this element
8. † required for POT calculation
9. § required for weighting (e.g. relocation calculation of "x-y weight")

2.2 general entry info

Table 2 details some basic elements. The `run` branch is repetitive within a file but useful to distinguishing entries when the `TTree`s are chained together. Prior to the addition of any metadata to the file, the range of `evtno` values was used make a calculated *guess* at the total protons-on-target (POTs) the file represents. Because not every proton generates an entry in the `TTree` and because for some formats in some cases the proton number was lost (e.g. muon decays in flugg) one can not simply use the difference in the first and last entries.

Variable		g3	flugg	g4	lbne	new	Description
`run`		I	I	I	I	I	Run number (arbitrary)
`evtno`	†	I	I	I	I	I	Event number (proton on target)

Table 2: General entry information.

2.3 fixed decays

Table 3 represents the results of decays where the neutrino ray direction is either chosen randomly or forced through a particular point. The random decay is just that: whatever GEANT4 (or whatever) generated. The other tuples are calculated by limiting the ray to going through a given point. This choice will affect the neutrino's energy and direction and will have an associated weight (probability).

For a "far" detector far enough away that subtends a small enough solid angle the choice of a single point is relatively insignificant as the beam is essentially a parallel plane wave front. But this is not true for any sizable "near" detector which will see a line source rather than a point source and thus is subject to variation in energy spectra and intensity throughout its volume. Thus the "near" values can not be used as-is in event generators such as GENIE if they are to represent a detailed simulation. They are adequate for some crude purposes to get a general feel for different locations.

One could condense this section down to simple vectors of **`ndxdz`**, **`ndydz`**, **`npz`**, **`nenergy`**, **`nwt`** where element [0] would represent the random decay (**`nwt=1`**), and subsequent elements hold some mixture of various "near" and "far" locations. This is something to consider; for now I've left the three cases as separate elements. Currently files lack any metadata that tells one what location a "near" or "far" entry represents. For instance **`flugg`** files might have MINOS or NOνA locations used depending on who generated the file; this has led to surprises for the unwary and additional headaches when trying to rectify the differences seen by people running essentially the same code.

Variable	g3	flugg	g4	lbne	new	Description
`Ndxdz` `Ndydz`	F	D	D	F	D	ν direction slopes for a random decay
`Npz`	F	D	D	F	D	ν momentum (GeV/c) along the z-axis (beam axis) for a random decay
`Nenergy`	F	D	D	F	D	ν energy (GeV) for a random decay
`NdxdzNear` `NdydzNear`	F	D	D[11]	F[5]	<D>	Direction slopes for a ν forced towards the center of the "near" detector(s)
`NenergyN`	F	D	D[11]	F[5]	<D>	Energy for a ν forced towards the center of the "near" detector(s)
`NWtNear`	F	D	D[11]	F[5]	<D>	Weight for a ν forced towards the center of the "near" detector(s)
`NdxdzFar` `NdxdzFar`	F	D	D[2]	F[3]	<D>	Direction slopes for a ν forced towards the center of the "far" detector(s)
`NenergyF`	F	D	D[2]	F[3]	<D>	ν energy (GeV) for a decay forced to the center of the "far" detector(s)
`NWtFar`	F	D	D[2]	F[3]	<D>	ν weight for a decay forced to the center of the far detector(s)

Table 3: Limited neutrino ray information.

2.4 decay data

Table 4 is (mostly) the core information about the neutrino and the decay that gave rise to it. From the information marked with a § one can calculate the energy and weight for the neutrino ray to go through any point (small angles assumed??).

Variable		g3	flugg	g4	lbne	new	Description
Norig		I	I	I	I	I	neutrino origin: `g4numi`: 1=particle from target (or baffle), 2=from scraping, 3=from μ decay (Not filled in `flugg`)
Ndecay	¶	I	I	I	I	I	Decay process that produced the ν, see Table 10
Ntype	§	I	I	I	I	I	ν flavor. ‡GEANT codes: $\nu_\mu = 56, \bar{\nu}_\mu = 55, \nu_e = 53, \bar{\nu}_e = 52$
Vx Vy Vz	§	F	D	D	F	D	ν production vertex (cm)
pdPx pdPy pdPz	§	F	D	D	F	D	Momentum (GeV/c) of the ν parent at the ν production vertex (parent decay point)
ppdxdz ppdydz	§	F	D	D	F	D	Direction of the ν parent at its production point (which may be in the target)
pppz	§	F	D	D	F	D	z momentum (GeV/c) of the ν parent at its production point
ppenergy	§	F	D	D	F	D	Energy (GeV) of the ν parent at its production point
ppmedium	¶	I	I	D	F	?	Code for the material the ν parent was produced in (see Table 10)
ptype	§	I	I	I	I	I	ν parent species (GEANT codes‡)
ptrkid		-	-	-	I	?	**need lbne description**
ppvx ppvy ppvz		F	D	D	F	D	Production vertex (cm) of the ν parent
muparpx muparpy muparpz	§	F	D	D	F	D	Momentum (GeV/c) of the ν grandparent at the grandparent decay point (muons) or grandparent production point (hadrons) (at the decay point in production files – see footnote on page **??**
mupare	§	F	D	D	F	D	Energy (GeV) of the ν grandparent, as above
Necm	§	F	D	D	F	D	ν energy (GeV) in the center-of-mass frame
Nimpwt	§	F	D	D	D	D	Importance weight of the ν

Table 4: The core information about the decays.

2.5 parent data

Entries marked with a ¶ represent info (beyond §) that MINOS or NOνA might use to in reweighting.

The `beamHWidth` through `hornCurrent` (and `protonN`) elements (found in the G4NUMI and G4LBNE layouts immediately after `evtno`) are presented here, out-of-order, because they seem related to others in this section. Most of those seem to be metadata (can anyone confirm this?) that won't vary from entry to entry. The `flugg`-only entries in Table 6 are derived values.

Variable	g3	flugg	g4	lbne	new	Description
xpoint ypoint zpoint	F	D	D	F	?	(Not filled in `flugg`, others?)
tvx tvy tvz	F	D	D	F	D	Position (cm) of the ν ancestor as it exits target (possibly, but not necessarily, the direct ν parent)
tpx tpy ¶ tpz	F	D	D	F	D	Momentum (GeV/c) of the ancestor as it exits target
tptype ¶	I	I	I	I	I	Species of the ancestor exiting the target (GEANT codes‡)
tgen	I	I	I	I	I	ν parent generation in cascade. 1 = primary proton, 2 = particles produced by proton interaction, 3 = particles from 2's
tgptype	I	I	-	-	?	Species of the parent of the particle exiting the target (GEANT codes‡)
tgppx tqppy tqppz	F	D	-	-	?	Momentum (GeV/c) of the parent of the particle exiting the target at the parent production point (at the decay point in production files – see footnote on page **??**
tprivx tprivy tprivz	F	D	-	-	?	Primary particle interaction vertex (not used)
beamx beamy beamz	F	D	-	-	?	Primary proton origin (cm)
beampx beampy beampz	F	D	-	-	?	Primary proton momentum (GeV/c)
protonN	-	-	-	I	?	**need lbne description of difference w/** `evtno`
beamHWidth beamVWidth	-	-	D	F	?	**need g4numi description**
beamX beamY	-	-	D	F	?	**need g4numi description**
protonX protonY protonZ	-	-	D	F	?	**need g4numi description**
protonPx protonPy protonPz	-	-	D	F	?	**need g4numi description**
nuTarZ	-	-	D	F	?	**need g4numi description**
hornCurrent	-	-	D	F	?	**need g4numi description**

Table 5: Miscellaneous information, mostly do to with some ancestors.

Variable	g3	flugg	g4	lbne	new	Description
Vr	-	D	-	-	?	$\sqrt{(\mathtt{Vx}^2 + \mathtt{Vy}^2)}$
pdP	-	D	-	-	?	$\sqrt{(\mathtt{pdPt}^2 + \mathtt{pdPz}^2)}$
pdPt	-	D	-	-	?	$\sqrt{(\mathtt{pdPx}^2 + \mathtt{pdPy}^2)}$
ppp	-	D	-	-	?	$\sqrt{(\mathtt{pppt}^2 + \mathtt{pppz}^2)}$
pppt	-	D	-	-	?	$\sqrt{(\mathtt{ppdxdz}^2 + \mathtt{ppdydz}^2)} \times \mathtt{pppz}$
ppvr	-	D	-	-	?	**filled with tvr calculation, should be:** $\sqrt{(\mathtt{ppvx}^2 + \mathtt{ppvy}^2)}$
muparp	-	D	-	-	?	$\sqrt{(\mathtt{muparpt}^2 + \mathtt{muparpz}^2)}$
muparpt	-	D	-	-	?	$\sqrt{(\mathtt{muparpx}^2 + \mathtt{muparpy}^2)}$
tvr	-	D	-	-	?	**never filled! looks like typo stores calculated value in ppvr, should be:** $\sqrt{(\mathtt{tvx}^2 + \mathtt{tvy}^2)}$
tp	-	D	-	-	?	$\sqrt{(\mathtt{tpt}^2 + \mathtt{tpz}^2)}$
tpt	-	D	-	-	?	$\sqrt{(\mathtt{tpx}^2 + \mathtt{tpy}^2)}$

Table 6: flugg helper variables.

2.6 ancestor data

Table 7 is primarily `g4numi` and MINERνA's additions. Leo/? should verify the descriptions. By using STL `vectors` rather than fixed sized arrays we can eliminate the need for `ntrajectory` and `overflow`. Most of these need tweaks to the name to identify them as being information about the intermediate particles. Questions

- what do trackId and ParentId represent?
- ivol? fvol?
- trkx vs. startx, trkpx vs. startpx?
- Isn't start*[n] = stop*[n-1] ?
- choice of `TString` vs. STL `string`?
- is entry [0] the proton?
- is entry [ntrajectory] the decaying particle?
- indications in code that some of these entries use mm and MeV as units, which is at odds with the units for other variables

Variable	g4	mnv	new	Description

`trkx`
`trky`
`trkz`
	`D[10]`	`D[10]`	<D>	??? Origin of intermediate **descriptive name?** *minerva check*

`trkpx`
`trkpy`
`trkpz`
	`D[10]`	`D[10]`	<D>	??? Momentum at origin of intermediate **descriptive name?** *minerva check*

`ntrajectory`	-	`I`	-	Number of intermediate levels *minerva check*
`overflow`	-	`B`	-	Flag list as incomplete *minerva check*
`pdg`	-	`I[10]`	<I>	Intermediate's particle type **descriptive name?**
`trackId`	-	`I[10]`	<I>	??? **descriptive name?**
`parentId`	-	`I[10]`	<I>	??? **descriptive name?**

`startx`
`starty`
`startz`
	-	`D[10]`	<D>	??? Origin of intermediate **descriptive name?** *minerva difference w/* `trk` *above*

`stopx`
`stopy`
`stopz`
	-	`D[10]`	<D>	??? End of intermediate **descriptive name?** *minerva check*

`startpx`
`startpy`
`startpz`
	-	`D[10]`	<D>	??? Momentum at origin of intermediate **descriptive name?** *minerva difference w/* `trk` *above*

`stoppx`
`stoppy`
`stoppz`
	-	`D[10]`	<D>	??? Momentum at end of intermediate **descriptive name?** *minerva check*

`pprodpx`
`pprodpy`
`pprodpz`
	-	`D[10]`	<D>	??? **descriptive name?** *minerva check*

`proc`	-	`TS[10]`	<s>	??? **descriptive name?**
`ivol`	-	`TS[10]`	<s>	??? **descriptive name?**
`fvol`	-	`TS[10]`	<s>	??? **descriptive name?**

Table 7: Information about intermediates between the proton and the decaying particle.

2.7 proposed primary ntuple additions and metadata

Table 8 suggests some possible addtions. By providing STL `vectors` of integers and doubles users can add data that they need, especially for temporary short term studies, without having to change the basic format – which would affect all other users. The mapping from index into the vector to *meaning* will necessarily be up to the user. For cases where every entry has the same fixed mapping we would provide name vectors in the metadata to record that ordering. If the sizes vary on an entry by entry basis then it is left to the user to keep it straight.

I am also proposing the addition of a `flagbits` branch. My initial thoughts on this were to allow single bits to signal information. Some bits would be reserved for fixed purposes and and the rest would be up for individual user designation. One idea here would be to reserve bits to flag choices for units (currently these are expected to be cm for length, GeV for energy & momentum, but the user might prefer meters or mm and MeV) and particle codes (currently expected to be GEANT3 with ν extensions, but it would be nice to uniformly use PDG codes by default). While these suggested bits would generally be of file-wide scope the additional cost of one integer per entry is minimal.

For the file-level metadata we need a name for the tree.

Variable	new	Description
`vint`	<I>	STL `vector` of integers, for users to fill as they please
`vdbl`	<D>	STL `vector` of doubles, for users to fill as they please
`flagbits` ‡	I	Flags to indicate units and particle numbering scheme; some bits reserved for user designation

Table 8: Proposed additions for the primary ntuple (i.e. one entry per decay).

Variable	new	Description
`simversion`	s	Name and version of program that generated file.
`pots`	D	Corresponding protons-on-target for the ntuple.

`xlocnear`
`ylocnear`
`zlocnear`
	<D>	Position info for each of the "near" locations.

`xlocfar`
`ylocfar`
`zlocfar`
	<D>	Position info for each of the "far" locations.

`vintnames`	<s>	STL `vector` of `strings` to hold names for `vint` elements.
`vdblnames`	<s>	STL `vector` of `strings` to hold names for `vdbl` elements.

Table 9: Proposed metadata elements (i.e. one entry per file).

3 Defining the `TTree`

The `gnumi` (GEANT3) ntuple is created using `hbook` as a column-wise (`common block`-based) ntuple. The ROOT version is generated by using `h2root` to convert it from the ZEBRA file format. As generation of new beamline simulations using this code is unlikely we will not further comment on the necessary steps for converting to the new format (it would be difficult).

3.1 `flugg`

The `flugg` `TTree` is filled using the script `numisoft/g4numi_flugg/root/fill_flux.C` which reads data from an ASCII text file. The extra ("extended") elements discussed in Table 6 are calculated when creating the entry; they are also apparently partially *kaput* (it's a technical term) due to a cut-and-paste typo.

```
...
  TFile *ft = new TFile(ftree,"recreate");
  TTree *mtree = new TTree("h10","neutrino");
  int    run;      mtree->Branch("run",      &run,      "run/I");     //1
  int    evtno;    mtree->Branch("evtno",    &evtno,    "evtno/I");   //2
...
  double Ndxdznea; mtree->Branch("Ndxdznea", &Ndxdznea, "Ndxdznea/D");//7
...
  int events = 0;
  while(!datafile.eof()) {
    // read a line from the text file
       datafile
           >> run      //1
           >> evtno    //2
           ...
           >> beampz ; //62
...
       mtree->Fill();
       ++events;
  }
  datafile.close();
  mtree->Write();
  ft->Close();
```

To make this work for the new file format basically involve changing the branch names, adding new branches and changing the types for those that are fixed sized arrays, making them vectors. Untested code follows:

```
#include <string>
#include <vector>
using namespace std;
...
  int bufsiz = 32000;  // best value?
  int splitlvl = 99;   // best value?
...
  std::vector<double> ndxdznear;
  mtree->Branch("ndxdznear","vector<double>",  &ndxdznear, bufsiz, splitlvl);
  ndxdznear.reserve(1);  // we know there will always be only one value (flugg files)
                         // and we must reserve space to have somewhere to put the value
                         // (this is less intensive than clear/push_back pairs in the loop)
...
  while(!datafile.eof()) {
    // read a line from the text file
...
       >> ndxdznear[0]  // already reserved space, so we can set it
...
```

Alternatively, with a minor reworking of the code the script could be rewritten to use compiled code and the actual structure. The would be the preferred route forward. The framework for this upgrade can be found in Section 5.

An inspection of this script (`numisoft/g4numi_flugg/root/fill_flux.C`) turned up an error that needs to be fixed and committed back to all repository instances. The error is an obvious cut-and-paste typo:

```
  if (extend) {
     Vr = SumSq(Vx, Vy);
     pdPt = SumSq(pdPx, pdPy);
     pdP = SumSq(pdPt, pdPz);
     pppt = SumSq(ppdxdz, ppdydz)*pppz;
     ppp = SumSq(pppt, pppz);
     ppvr = SumSq(ppvx, ppvy);
     muparpt = SumSq(muparpx, muparpy);
     muparp = SumSq(muparpt, muparpz);
     ppvr = SumSq(tvx, tvy);  // the left hand side of this assignment should be "tvr"
                              // and not a repeat of "ppvr"
     tpt = SumSq(tpx, tpy);
     tp = SumSq(tpt, tpz);
  }
```

3.2 g4numi and variants

The `g4numi` `TTree` is filled in compiled code in `numisoft/g4numi/src/NumiAnalysis.cc`. The basic `TTree` is simply the series of `data_t` class objects, and is booked and filled via:

```
NumiAnalysis::NumiAnalysis()
...
    // individual entries in the tree are "data_t" objects
    g4data = new data_t();  // this is a private data member

void NumiAnalysis::book()
...
    nuNtuple = new TFile(nuNtupleFileName,"RECREATE","root ntuple");
    tree = new TTree("nudata","g4numi Neutrino ntuple");
    tree->Branch("data","data_t",&g4data,32000,1);

void NumiAnalysis::FillNeutrinoNtuple(const G4Track& ...
...
     // set values in g4data
     g4data->run = ...
     ...// loop for elements that are arrays
       g4data->NdxdzNear[ii] = ...
     ...
     tree->Fill();

void NumiAnalysis::finish()
...
nuNtuple->cd();
tree->Write();
nuNtuple->Close();
delete nuNtuple;
```

A couple of issues, as currently implemented, with this approach that I've noticed include:

1. the version number in the `data_t.hh` have never been incremented even when the layout changes (i.e. `ClassDef(data_t,1)` in `data_t.hh` always). In the new scheme one needs to always be sure to increment the version number whenever the data layout changes.

2. `g4data->Clear()` is never called, which means that entries that that vary in length (i.e. most of the MINERνA additions) retain high water values beyond the current `ntrajectory` from previous entries. This isn't an issue if one never indexes into the array beyond the current entry's set of values, but it can be confusing and it will cause the file to be larger than necessary (random values don't compress as well as 0).

The new ntuple format would be simply replacing the `data_t` with a new class. Member variable names would need adjustments in the `NumiAnalysis` code. Additionally, one would want to apply the `Clear()` method before the fill, which should reset any STL `vectors` to have zero length. Any instances of using fixed indexing during filling would need to be converted to `push_back()` methods on the element, i.e.:

```
//OLD: g4data->NdxdzNear[ii] = ...
dk2nu->ndxdznear.push_back(...);
```

4 Proposal

```
/**
 * \class dk2nu
 * \file  dk2nu.h
 *
 * \brief A class that defines the "dk2nu" object used as the primary
 *        branch for a TTree for the output of neutrino flux simulations
 *        such as g4numi, g4numi_flugg, etc.
 *
 * \author (last to touch it) $Author: rhatcher $
 *
 * \version $Revision: 1.1 $
 *
 * \date $Date: 2012/04/02 21:19:46 $
 *
 * Contact: rhatcher@fnal.gov
 *
 * $Id: dk2nu.h,v 1.1 2012/04/02 21:19:46 rhatcher Exp $
 *
 * Notes tagged with "DK2NU" are questions that should be answered
 */

#ifndef DK2NU_H
#define DK2NU_H

#include "TROOT.h"
#include "TObject.h"

#include <vector>
#include <string>

class dk2nu
{
private:
  ClassDef(dk2nu,1) // KEEP THIS UP-TO-DATE!  increment for each change

public:
  /**
   *   Public methods for constructing/destruction and resetting the data
   */
  dk2nu();
  virtual ~dk2nu();
  void Clear(const std::string &opt = ""); ///< reset everything to undefined

  /**
   *  All the data members are public as this class is used as a
   *  generalized struct, with just the addition of the Clear() method.
   *  As they will be branches of a TTree no specialized naming
   *  indicators signifying that they are member data of a class
   *  will be used, nor will any fancy capitalization schemes.
   */

  /**
```

```
   *======================================================================
   *  General Info
   */
   Int_t run;              ///< identifying run #
   Int_t evtno;            ///< proton # processed by simulation

  /**
   *======================================================================
   *  Fixed Decays:
   *  A random ray plus those directed at specific points.
   */
   Double_t ndxdz;        ///< dx/dz direction slope for random decay
   Double_t ndydz;        ///< dy/dz direction slope for random decay
   Double_t npz;          ///< z-axis momentum for random decay
   Double_t nenergy;      ///< neutrino energy for random decay

   std::vector<Double_t> ndxdznear;  ///< dx/dz slope for near detector(s)
   std::vector<Double_t> ndydznear;  ///< dy/dz slope for near detector(s)
   ///< DK2NU: add npznear ?
   std::vector<Double_t> nenergyn;   ///< energy for near detector(s)
   std::vector<Double_t> nwtnear;    ///< weight for near detector(s)

   std::vector<Double_t> ndxdzfar;   ///< dx/dz slope for near detector(s)
   std::vector<Double_t> ndydzfar;   ///< dy/dz slope for near detector(s)
   ///< DK2NU: add npzfar ?
   std::vector<Double_t> nenergyf;   ///< energy for near detector(s)
   std::vector<Double_t> nwtfar;     ///< weight for near detector(s)

  /**
   *======================================================================
   *  Decay Data:
   *  Core information about the neutrino and the decay that gave rise to it.
   *  % = necessary for reweighting
   */
   Int_t    norig;        ///< not used?
   Int_t    ndecay;       ///< decay process (see dkproc_t)
   Int_t    ntype;        ///< % neutrino flavor (PDG? code)

   Double_t vx;           ///< % neutrino production vertex x
   Double_t vy;           ///< % neutrino production vertex y
   Double_t vz;           ///< % neutrino production vertex z
   Double_t pdpx;         ///< % px momentum of nu parent at (vx,vy,vz)
   Double_t pdpy;         ///< % py momentum of nu parent at (vx,vy,vz)
   Double_t pdpz;         ///< % pz momentum of nu parent at (vx,vy,vz)

   /**  these are used in muon decay case? */
   Double_t ppdxdz;       ///< % direction of nu parent at its production point
   Double_t ppdydz;       ///< % direction of nu parent at its production point
   Double_t pppz;         ///< % z momentum of nu parent at its production point
   Double_t ppenergy;     ///< % energy of nu parent at its production point

   Double_t ppmedium;     ///< material nu parent was produced in
   Int_t    ptype;        ///< % nu parent species (PDG? code)

```

```
     /** momentum and energy of nu grandparent at
         muons:    grandparent decay point
         hadrons:  grandparent production point
         Huh?  this needs better documentation
      */
     Double_t muparpx;      ///< %
     Double_t muparpy;      ///< %
     Double_t muparpz;      ///< %
     Double_t mupare;       ///< % energy of nu grandparent

     Double_t necm;         ///< % nu energy in center-of-mass frame
     Double_t nimpwt;       ///< % production vertex z of nu parent

    /**
     *=====================================================================
     *  (Grand)Parent Info:
     *
     */

     /**
      * DK2NU: are these needed for any/all cases?
      */
     Double_t ppvx;         ///< production vertex x of nu parent
     Double_t ppvy;         ///< production vertex y of nu parent
     Double_t ppvz;         ///< production vertex z of nu parent

     /**
      * DK2NU: do we need these?  these aren't filled by flugg, others?
      */
     Double_t xpoint;       ///< ?
     Double_t ypoint;       ///< ?
     Double_t zpoint;       ///< ?

     /**
      * these ancestors are possibly, but not necessarily, the direct nu parent
      * DK2NU: can these be removed in favor of cascade info below?
      */
     Double_t tvx;          ///< x position of nu ancestor as it exits target
     Double_t tvy;          ///< y position of nu ancestor as it exits target
     Double_t tvz;          ///< z position of nu ancestor as it exits target
     Double_t tpx;          ///< x momentum of nu ancestor as it exits target
     Double_t tpy;          ///< y momentum of nu ancestor as it exits target
     Double_t tpz;          ///< z momentum of nu ancestor as it exits target
     Int_t    tptype;       ///< species of ancestor exiting the target
     Int_t    tgen;         ///< nu parent generation in cascade:
                            ///<   1=primary proton
                            ///<   2=particles produced by proton interaction
                            ///<   etc
     /**
      * these are only in g3numi and flugg
      * DK2NU: can these be removed in favor of cascade info below?
      *        for now we'll leave them in place
      */
     Int_t    tgptype;      ///< species of parent of particle exiting the target (PDG code?)
```

```

    Double_t tgppx;        ///< x momentum of parent of particle exiting target at the parent prod
    Double_t tgppy;        ///< y momentum
    Double_t tgppz;        ///< z momentum
    Double_t tprivx;       ///< primary particle interaction vtx (not used?)
    Double_t tprivy;       ///< primary particle interaction vtx (not used?)
    Double_t tprivz;       ///< primary particle intereaction vtx (not used?)
    Double_t beamx;        ///< primary proton origin
    Double_t beamy;        ///< primary proton origin
    Double_t beamz;        ///< primary proton origin
    Double_t beampx;       ///< primary proton momentum
    Double_t beampy;       ///< primary proton momentum
    Double_t beampz;       ///< primary proton momentum

    /**
     * these are in the g4numi and minerva ntuples
     * DK2NU: but what do they mean and are the duplicative to
     *        the more complete progenitor info below?
     */
    std::vector<Double_t> trkx;
    std::vector<Double_t> trky;
    std::vector<Double_t> trkz;
    std::vector<Double_t> trkpx;
    std::vector<Double_t> trkpy;
    std::vector<Double_t> trkpz;

   /**
    *=========================================================================
    *  Progenitor Info:
    *  Complete ancestral info from primary proton down to decaying particle
    *
    *  DK2NU: this is mainly (based on) the minerva extensions *except*
    *         some names are changed to avoid confusion and
    *         distances will be cm, energies in GeV (unless the whole
    *         record uniformly uses something else and is flagged as such)
    */
    std::vector<Int_t>    apdg;     ///< ancestor species
    std::vector<Int_t>    trackid;  ///< ??? particle trackId
    std::vector<Int_t>    parentid; ///< ??? parentId

    std::vector<Double_t> startx;   ///< particle x initial position
    std::vector<Double_t> starty;   ///< particle y initial position
    std::vector<Double_t> startz;   ///< particle z initial position
    std::vector<Double_t> stopx;    ///< particle x final position
    std::vector<Double_t> stopy;    ///< particle y final position
    std::vector<Double_t> stopz;    ///< particle z final position

    std::vector<Double_t> startpx;  ///< particle x initial momentum
    std::vector<Double_t> startpy;  ///< particle y initial momentum
    std::vector<Double_t> startpz;  ///< particle z initial momentum
    std::vector<Double_t> stoppx;   ///< particle x final momentum
    std::vector<Double_t> stoppy;   ///< particle y final momentum
    std::vector<Double_t> stoppz;   ///< particle z final momentum

```

```
  std::vector<Double_t> pprodpx;  ///< parent x momentum when producing this particle, MeV/c
  std::vector<Double_t> pprodpy;  ///< parent y momentum when producing this particle
  std::vector<Double_t> pprodpz;  ///< parent z momentum when producing this particle

  std::vector<std::string> proc;  ///< name of the process that creates this particle

  std::vector<std::string> ivol;  ///< name of the volume where the particle starts
  std::vector<std::string> fvol;  ///< name of the volume where the particle stops

  /**
   *========================================================================
   *  Special Info:
   */
  Int_t    flagbits;      ///< bits signify non-std setting such as
                          ///< Geant vs. PDG codes, mm vs. cm, Mev vs. GeV
  std::vector<Int_t>    vint;    ///< user defined vector of integers
  std::vector<Double_t> vdbl;    ///< user defined vector of doubles

  /**
   *========================================================================
   *  Random Info:
   *  blah, blah, blah
   */

  Int_t    ptrkid;        ///< lbne addition

  /**
   * stuff that should be in the metadata
   */

  /**
   *========================================================================
   *  Specialized enumerations
   */

  /**
   *  Proposed flag bits:
   */
  typedef enum flgbitval {
    flg_dist_m          = 0x00000000,  ///< no special bit for meters
    flg_dist_cm         = 0x00020000,  ///< distances in cm (default)
    flg_dist_mm         = 0x00030000,  ///< distances in mm
    flg_e_gev           = 0x00000000,  ///< no special bit for GeV (default)
    flg_e_mev           = 0x00300000,  ///< energies in MeV
    flg_usr_mask        = 0x0000FFFF,
    flg_reserved_mask   = 0xFFFF0000
  } flgbitval_t;

  /**
   *  Enumeration of decay processes, stored in "ndecay"
   *  store as integer; these are for reference
   *  DK2NU:  should there be an associated AsString() method
   *          that returns a text (optionally formatted for latex?)?
   */
```

```
    typedef enum dkproc {
      dkp_unknown          =  0,
      dkp_k0l_nuepimep     =  1,  ///< k0long => nu_e + pi- + e+
      dkp_k0l_nuebpipem    =  2,  ///< k0long => nu_e_bar + p+ + e-
      dkp_k0l_numupimmup   =  3,  ///< k0long => nu_mu + pi- + mu+
      dkp_k0l_numubpipmum  =  4,  ///< k0long => nu_mu_bar + pi+ + mu-
      dkp_kp_numumup       =  5,  ///< k+ => nu_mu + mu+
      dkp_kp_nuepi0ep      =  6,  ///< k+ => nu_e + pi0 + e+
      dkp_kp_numupi0mup    =  7,  ///< k+ => nu_mu + pi0 + mu+
      dkp_kp_numubmum      =  8,  ///< k- => nu_mu_bar + mu-
      dkp_kp_nuebpi0em     =  9,  ///< k- => nu_e_bar + pi0 + e-
      dkp_kp_numubpi0mum   = 10,  ///< k- => nu_mu_bar + pi0 + mu-
      dkp_mup_nusep        = 11,  ///< mu+ => nu_mu_bar + nu_e + e+
      dkp_mum_nusep        = 12,  ///< mu- => nu_mu + nu_e_bar + e-
      dk_pip_numumup       = 13,  ///< pi+ => nu_mu + mu+
      dk_pim_numubmum      = 14,  ///< pi- => nu_mu_bar + mu-
      dkp_maximum,                ///< one-beyond end for iterating
      dkp_other            = 999, ///< flag for unusual cases
    } dkproc_t;

  };

#endif
```

5 Example test program for filling

```
//
// test creating and filling a TTree based on dk2nu.h (dk2nu.C)
// this script can be run using:
//       root -b -q test_fill_dk2nu.C+
//
// rhatcher@fnal.gov  2012-04-03
//====================================================================

#include "dk2nu.h"

// include this because we're not linking to anything external
// so we need to include the source for dk2nu::Clear()
#include "dk2nu.cc"

// make a dictionary for dk2nu class, again because no external linkages
#ifdef __CINT__
#pragma link C++ class dk2nu+;
#endif

#include "TFile.h"
#include "TTree.h"
#include "TRandom3.h"

// 510000 seems to be an upper limit on # of entries for flugg 500K POT lowth
void test_fill_dk2nu(unsigned int nentries=1000)
{

```

```
  // stuff...
  TRandom3* rndm = new TRandom3();

  // create object
  dk2nu* dk2nuObj = new dk2nu;

  // create file, book tree, set branch address to created object
  TFile* treeFile = new TFile("test_dk2nu.root","RECREATE");
  TTree* tree     = new TTree("dk2nu","FNAL neutrino ntuple");
  tree->Branch("dk2nu","dk2nu",&dk2nuObj,32000,1);

  // fill a few element of a few entries
  for (unsigned int ipot=1; ipot <= nentries; ++ipot) {
    // clear the object in preparation for filling an entry
    dk2nuObj->Clear();

    // fill with info ... only a few elements, just for test purposes
    dk2nuObj->run   = 42;
    dk2nuObj->evtno = ipot;
    // just test the filling of vector
    unsigned int nancestors = rndm->Integer(12) + 1;  // at least one entry
    for (unsigned int janc = 0; janc < nancestors; ++janc ) {
      int xpdg = rndm->Integer(100);
      dk2nuObj->apdg.push_back(janc*10000+xpdg);
    }

    // push entry out to tree
    tree->Fill();

  } // end of fill loop

  // finish and clean-up
  treeFile->cd();
  tree->Write();
  treeFile->Close();
  delete treeFile; treeFile=0; tree=0;
}
```

6 Example use of the tree in a ROOT session

```
TFile* myfile = TFile::Open("test_dk2nu.root","READONLY");
TTree* mytree = 0;
myfile->GetObject("dk2nu",mytree);
mytree->Scan("run:evtno:@apdg.size():apdg[2]");
```

The `@` in `@apdg.size()` is the ROOT mechanism for signaling that the `.size()` method is to be applied to the collection as a whole and not on individual items, so this prints the length of the `apdg` STL `vector`. The `apdg[2]` prints the 3rd entry (if it exists); using `[]` (or giving none) for vectors performs an implicit loop. The looping rules for `Scan()` or `Draw()` on array elements in `TTree`s are complex and appropriate documentation should be consulted[1].

[1] http://root.cern.ch/root/html/TTree.html#TTree:Draw@2

7 Auxillary numbering schemes

Ndecay	Process
1	$K_L^0 \to \nu_e + \pi^- + e^+$
2	$K_L^0 \to \bar{\nu}_e + \pi^+ + e^-$
3	$K_L^0 \to \nu_\mu + \pi^- + \mu^+$
4	$K_L^0 \to \bar{\nu}_\mu + \pi^+ + \mu^-$
5	$K^+ \to \nu_\mu + \mu^+$
6	$K^+ \to \nu_e + \pi^0 + e^+$
7	$K^+ \to \nu_\mu + \pi^0 + \mu^+$
8	$K^- \to \bar{\nu}_\mu + \mu^-$
9	$K^- \to \bar{\nu}_e + \pi^0 + e^-$
10	$K^- \to \bar{\nu}_\mu + \pi^0 + \mu^-$
11	$\mu^+ \to \bar{\nu}_\mu + \nu_e + e^+$
12	$\mu^- \to \nu + \bar{\nu}_e + e^-$
13	$\pi^+ \to \nu_\mu + \mu^+$
14	$\pi^- \to \bar{\nu}_\mu + \mu^-$
999	Other

Code	Material
5	Beryllium
6	Carbon
9	Aluminum
10	Iron
11	Slab Steel
12	Blu Steel
15	Air
16	Vacuum
17	Concrete
18	Target
19	Rebar Concrete
20	Shotcrete
21	Variable Density Aluminum
22	Variable Density Steel
23	1018 Steel
24	A500 Steel
25	Water
26	M1018 Steel
28	Decay Pipe Vacuum
31	CT852

Table 10: The decay codes stored in `ndecay` and material codes as defined by Gnumi and used in the fluxfiles, old and current.